DOCUMENT RESUME

ED 126 140 TM 005 396

AUTHOR Marsh, Herbert W.; And Others
TITLE The Relationship Between Students' Evaluation of Instruction and Expected Grades.
PUB DATE [Apr 76]
NOTE 26p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Bias; *College Students; College Teachers; *Correlation; Course Evaluation; *Expectation; Factor Analysis; *Grades (Scholastic); Higher Education; Literature Reviews; *Student Evaluation of Teacher Performance; Teacher Characteristics; Validity

ABSTRACT

The relationship between grades that students expected to receive and their evaluations of instructional quality was investigated. Correlations between expected grades and 10 evaluation scores--eight evaluation factors and two overall summary items--were based on the average responses in 591 undergraduate classes offered one term at the University of California, Los Angeles. Average responses to the overall instructor and overall course items, items most often used to obtain a summary impression, showed statistically significant correlations with average expected grades even though factors most closely associated with teaching (Instructor Enthusiasm, Breadth of Coverage, Interaction, and Organization) showed much smaller relationships with expected grades. This suggests that the overall summary items are probably more subject to response biases than factor scores that are weighted averages of responses to more specific items. The magnitude of correlation between expected grades and evaluations reported in this study is similar to that reported in other studies, but is higher than is generally reported in literature reviews advocating the use of students' evaluations. It was concluded that even if grading leniency does produce a bias in students' evaluations--and this is only one possible explanation--the biases are relatively small and will not cause a poor instructor to be evaluated highly or a superior instructor to be evaluated poorly. (Author/RC)

**********
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
**********

AERA
WEDNESDAY SYMPOSIUM
DIVISION A
BALLROOM 5 (HILTON)
8⁴⁵–10¹⁵

ED126140

# The Relationship Between Students' Evaluations of Instruction and Expected Grades

Herbert W. Marsh,[1] J. U. Overall, and Christopher S. Thomas

University of California, Los Angeles

Running Head: Expected Grades

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM005 396

ABSTRACT

This study was undertaken to establish the relationship between grades that students expected to receive and their evaluations of instructional quality. Correlations between expected grades and 10 evaluation scores--8 evaluation factors and two overall summary items--were based upon the average responses in 591 undergraduate classes offered one term at the University of California, Los Angeles. Average responses to the overall instructor and overall course items, items most often used to obtain a summary impression, showed statistically significant correlations with average expected grades ($r = .32$ and $.38$, respectively), even though factors most closely associated with teaching (Instructor Enthusiasm, Breadth of Coverage, Interaction, and Organization) showed much smaller relationships with expected grades. This suggests that the overall summary items are probably more subject to response biases than factor scores that are weighted averages of responses to more specific items. The magnitude of correlation between expected grades and evaluations reported in this study is similar to that reported in a number of other studies, but is higher than is generally reported in literature reviews advocating the use of students' evaluations. Based on the evidence presented, however, it was concluded that even if grading leniency does produce a bias in students' evaluations--and this is only one possible explanation--the biases are relatively small and will not cause a poor instructor to be evaluated highly or a superior instructor to be evaluated poorly.

## The Relationship Between Students' Evaluations of Instruction and Expected Grades

There is often the suspicion or fear that variables unrelated to the quality of instruction will affect students' evaluations of instruction. The harshest critics of students' evaluations even suggest that an instructor need only give high grades and demand little work of students to receive high evaluations. The purpose of this paper is to investigate the relationship between students' evaluations and the grades that students expect to receive. A grading leniency bias--students giving higher (or lower) evaluations in expectation of receiving higher (or lower) grades--is one bias that, if established, could undermine confidence in the evaluation process.

The relationship between students' evaluations and course grades that students expect to receive is a complex issue. A positive relationship, under different circumstances, can either offer strong support for the validity of students' evaluations or argue for a dangerous bias in their application. If higher grades received by students are indicative of superior learning resulting from superior instruction--a goal of all teachers--the corresponding higher evaluations support the validity of the student's evaluations. However, if higher grades are only indicative of greater leniency in assigning grades, then any improved evaluations based upon the expectation of higher grades suggest a bias and undermine the validity of the students' evaluations.

Evidence for the validity of students' evaluations has been presented by Marsh, Fleiner and Thomas (1975). The average student evaluations for each section of a multi-section course correlated positively with student performance on a standardized

final examination. The sections did not differ in initial ability and the student evaluations were made before the final examination was taken or final grades were awarded. Students in any one section did not know how the average performance of their section compared with the average performance of other sections. Consequently the average grade expected by each section at the time of the evaluations, as opposed to actual performance on the subsequent examination, showed no difference across the different sections. Since alternative explanations were eliminated, this study supports the validity of students' evaluations. Other studies have reported similar findings (Elliot, 1950; Morsh, Burgess and Smith, 1956; Cohen and Berger, 1970; Frey, 1973; Doyle and Whitely, 1974).

However, the finding that student evaluations reflect superior learning does not rule out the possibility that the evaluations are also biased by grading leniency. A given class of students may receive higher grades because they learned more, because the instructor was an easy grader, or a combination of the two. Across any wide sample of classes the two possibilities are confounded. The existence of a grading leniency bias in students' evaluations can only be disproved if the correlation between expected grades and evaluations is low or nonexistent. Before reviewing the appropriate literature, several methodological issues will be considered.

Methodological Considerations

The first methodological issue is the temptation to imply causation from correlation. Virtually all emperical data describing the relationship between students' evaluations and other variables is correlational and any causal inferences drawn from

this data are very speculative at best.

The second methodological issue concerns the distinction between statistical significance and practical significance. Any test statistic based upon a large sample size may be significant from a statistical point of view, yet be so small as to be of little practical significance. For example, a correlation of $r = .16$ based upon a sample size of 600 is highly significant ($p<.001$), but accounts for only about 2.5% of the variance. All research describing the relationship between student evaluations and other variables must consider both statistical and practical significance.

— A crucial and less obvious methodological issue is the choice of the appropriate unit of analysis--the individual student's evaluation or the average evaluation given by all the students in an entire class. Should the relationship between expected grades and students' evaluations be determined by the correlation between the average grades expected by entire classes and the average evaluations given by those classes, or by the correlation between grades expected by individual student and the evaluations given by the individual students? Although both approaches have been used, there are several reasons that argue for the superiority of the class average as the appropriate unit of analysis.

When the relationship between students' evaluations and expected grades is based on individual student responses, the relationship must be determined from responses across many different classes. This relationship, when determined within a single class, is irrelevant to the question of whether or not grading leniency

biases students' evaluations. Grading leniency is a characteristic of the instructor that will affect all the students in a given class. The most able student may expect to receive a higher grade, but the grade is not necessarily a more "lenient" grade. In fact, a lenient grader will tend to give more lenient grades to the least able students--"A" students will get their A's, but"D"and"F"students will get B's and C's. Within a single class the relationship between grades and evaluations probably depends on the focus of the class. Both easy and difficult graders can expect a positive correlation if the class is directed towards the most able students--better evaluations by the better student--and a negative correlation if the class is directed towards the less able students. Correlations within single classes, even when computed within a number of different single classes, cannot be used to argue for or against a grading leniency bias, and the appropriate comparison is between the evaluations of different instructors who differ in respect to grading leniency.

Even when the relationship between expected grades and students' evaluations is determined across a number of different classes, practical considerations argue for the superiority of the class average as the appropriate unit of analysis. When students' evaluations are used for administrative decisions, for feedback to individual faculty, or for course selection by students, the results are almost always presented in terms of class averages. Thus the relevant question is whether or not these averages are biased by grading leniency. In particular, grading leniency can most appropriately be assessed at the class level. An individual student who is scholastically superior may expect to get an "A"

even when the teacher is a hard grader, but if every student in an entire class expects to receive an "A", then there is good reason to suspect that the instructor is an easy grader. All the students expecting to get A's may or may not be in for a surprise when they actually receive their grades, but it is their expectations rather than reality that may bias their evaluations.

Statistical, as well as practical, considerations argue for the superiority of the class average as the appropriate unit of analysis. Statistical tests used to describe the relationship are based upon the assumption that each unit of analysis is independent, and this is certainly not the case when many individual students judge the same teacher in the same class; responses from 100 different students evaluating 100 different instructors would be independent, but responses from 100 different students evaluating the same instructor would not. Finally, there is the problem of response reliability. The average evaluations based upon 20 or more individual responses are extremely reliable, but the individual student responses are not. Using the Spearman-Brown correction factor to estimate reliabilities based upon different sample sizes, an item that had a reliability of .9 for a sample of 20 responses would only have a reliability of about .3 for a sample size of one. The unreliability in both the student evaluations and judgments of expected grades tends to mask the true relationship between the two variables.

In summary, the relationships between expected grades and students' evaluations are generally based upon correlational data and any causal inferences are very speculative. Statistically

significant relationships, particularly when based upon large sample sizes, may be so small as to be of little practical significance, and so both practical and statistical significance should be considered. The relationship between expected grades and students' evaluations should be based upon responses across a large number of different classes and not responses within separate classes. Both practical and statistical considerations argue for the superiority of the average responses given by an entire class of students as the appropriate unit of analysis.

## Literature Review

Comprehensive literature reviews have typically reported that the correlation between students' evaluations and expected grades tends to be very low (Remmers, 1963; Costin, Greenough, and Menges, 1971; Hildebrand, Wilson and Dienst, 1971; McKeachie, 1973). However, a number of individual studies reporting substantial relationships between the two suggest that this generalization needs to be explored further.

Individual Student Responses. The pioneering research on students' evaluations at Purdue University (Remmers, 1928; 1930) has often been misinterpreted as providing evidence against any grading leniency bias. Remmers did find that there was no systematic relationship between scholastic achievement and students' evaluations--some instructors were rated more highly by their best students, others by their worst--but Remmers did not use letter grades as a measure of scholastic achievement. Instead, students, based upon information provided by instructors at the time of the evaluations, merely indicated whether or not they were in the top half of the class being evaluated. This ingenious measure of

scholastic achievement avoids any confusion introduced by different grading standards and does not even depend upon grading leniency. Even if expected grades were used, the studies were based upon correlations within separate classes, which as indicated previously, cannot be used to argue for or against any grading leniency bias.

Studies that have looked at the relationship between individual students' evaluations and expected grades across a number of different classes have generally found low to moderate positive relationships (Staraak, 1934; Voeks and French, 1960; Stewart and Malpass, 1966; Kooker, 1968; Caffrey, 1969; Weigel, Oetting and Tasto, 1971; Hildebrand, et. al., 1971; Bausell and Magoon, 1972; Granzin and Painter, 1973). A selected summary of studies using larger sample sizes provides some meaningful generalizations. Hildebrand, et. al. (1971) collected evaulations from all students in 51 classes previously identified as being taught by the best and worst teachers at a particular university. The correlation between the overall instructor evaluation and expected grade was $r = .09$ ($n = 1015$ students); expected grades accounted for less than 1% of the variance in overall instructor rating. Granzin and Painter (1973) collected evaluations from 17 different classes and found correlations between expected grade and ratings of the "overall course", "course content" and "instructor" of .21, .12 and .16 respectively ($n = 639$ students); expected grades accounted for between 2.5% and 4.4% of the variance in the three ratings. Kooker (1968) asked students to rate one instructor they had the previous term, and compared the evaluations of students who received A's, B's and C's for upper division and lower division students separately. Significant F-ratios were reported, the

grade accounting for 10% of the variance in overall ratings for upperclassmen and about 13% of the variance for lower classmen. Bausell and Magoon (1972), drawing from a sample of 17,000 evaluations, randomly selected groups of 500 students who expected to receive A's, B's, C's and D's. F-ratios computed for each of 29 evaluation items were all significant. Expected grades accounted for about 14% of the overall course evaluation, 7% of the overall instructor evaluation, and an average of about 5% across all 29 items. In addition to the overall rating items, expected grade seemed most related to items referring to difficulty/workload and grading/examinations. However, somewhat lower relationships would probably have been found if the authors had used the same proportion of students expecting to receive each grade as had appeared in the population from which they were drawn.

In summary, studies considering the relationship between individual students' evaluations and expected grades have reported low to moderate relationships. These relationships are usually statistically significant when based upon sufficiently large sample sizes, but expected grades generally accounted for less than 10% of the variance in the evaluation items. The relationships tended to be higher for overall summary rating items, and particularly for the overall course rating.

Class Average Responses. Several experimenters have considered the relationship between average expected grade for an entire class and average evaluations given by the class. As previously indicated, this is the more appropriate unit of analysis for studying the effect of expected grades. Roshenshine, Cohen and Furst (1973) correlated average expected grade with average ratings of the instructor and

course,and found correlations of r = .09 and .27 respectively (n>1000 classes); expected grades accounted for about 1% and 7% respectively of the variance in the two items. Jiobu and Pollis (1971) reported that expected grade correlated .30 with overall course evaluation and .18 with amount of "Student Perceived Learning" (n = 67 classes); expected grades accounted for 9% and 3% respectively of the variance in the two items. Perry and Bauman (1973) found a correlation of r = .42(n = 123 classes) between average expected grade and the overall instructor rating; expected grades accounted for about 18% of the variance in this item. The relationship was somewhat higher for upper division courses than lower division courses. Anikeef (1953) correlated average grades actually obtained with average evaluations. Each average evaluation of an instructor was based upon the mean of 50 or more students' evaluations from at least 3 different classes. Across all course levels, the correlation was r = .51 (n = 19 instructors), but was higher for freshman-sophmore classes than it was for junior-senior classes. Note that while grades actually received accounted for 26% of the variance in evaluations, the small sample size suggests that this estimate is not very reliable.

In summary, studies considering the relationship between class average evaluations and class average expected grades have found moderate relationships. Expected grades generally accounted for close to 10% or more of the variance in at least one overall summary items. The magnitude of the relationship tends to be higher than was reported in studies based upon individual student responses.

Other Approaches. A rather unique approach to the study of the relationship between expected grades and students' evaluations was undertaken by Holmes (1972), who experimentally manipulated the effect of expected grades. Students in an introductory psychology course expecting to receive an A or a B were either given that grade or were given one grade lower than they earned. Holmes found that students who were given a grade lower than they earned subsequently evaluated the course significantly lower on 5 of 19 evaluation items (instructor preparation, did instructor have sufficient evidence to evaluate achievement, did you get less than expected from the course, and clarity of exam questions). However, even for items that showed statistically significant differences, the experimental manipulation accounted for less than 8% of the variation in any of evaluation items.

## Methods

### The Evaluation Instrument[2]

The evaluation instrument used in this study was developed while the first author was Director of the Evaluation of Instruction Program at the University of California, Los Angeles. Originally conceived as a means to improve undergraduate instruction, the instrument was designed to fulfill a host of objectives: sound statistical properties, practicality of usage, and acceptability by both students and faculty. Items with low reliabilities were eliminated; the median reliability for a class size of 25 is about .90. Items that faculty indicated as most useful and students indicated as most important were retained. Finally, factor analysis

was used extensively to find a reasonsable number of items that would adequately define distinct components of students' evaluations.

The research presented in this paper is based upon 21,000 evaluation forms completed by students in 591 undergraduate classes each with an enrollment of at least 10 students. The evaluations were conducted during the fall term in 1973 at the University of California, Los Angeles. A principle components factor analysis followed by a direct oblimin rotation (Harmon, 1968; Dixon, 1973) was performed on the class average evaluations. Eight evaluation factors were defined with sufficient clarity so that each individual item loaded higher on the factor it was designed to measure than on any other factor. The median intercorrelation between the weighted factor scores was r = .26. The difficulty factor tended to have low negative correlations with the other factors, while the remaining factors had low to moderate positive correlations with each other.

The eight evaluation factors and two overall summary items are:

Instructor Enthusiasm-- The instructor's display of enthusiasm, energy and ability to hold student interest while making valuable presentations.

Breadth-- The presentation of a broad background encompassing alternative approaches to the subject.

Organization-- The organization of the course, course materials, and class presentations.

Interaction-- The freedom students felt in interacting with the instructor and the value of these interactions.

Learning-- The extent to which students encounted a valuable learning experience.

Examinations-- Student perceptions of the value and fairness of graded materials in the course.

Assignments-- The value of class assignments (readings, homework, etc.) to the course.

Difficulty-- The relative difficulty and workload of the course and the pace of presentations.

Overall Instructor-- A single evaluation item asking "What is your overall rating of the instructor."

Overall Course-- A single evaluation item asking "What is your overall rating of the course."

Procedures

The evaluation forms were completed by students during the last two weeks of the 1973 fall quarter at UCLA. The actual mechanics of administering the forms varied for different academic departments. Generally the forms were distributed by the instructor, completed anonymously by students, placed in a large manilla envelope and immediately returned to the department coordinator by either the instructor or a student in the class. Students were informed that the results would be used for administrative decisions, and feedback to the faculty; also, results would be made publically available (with instructor permission) for use in student course selection. Instructors were not given the results of the evaluations until final grades had been assigned. The use of the particular form was not mandatory, but was used by most of the academic departments. Individual instructors were generally urged to use the evaluation instrument by department chairmen, but actual participation was voluntary.

Statistical Analysis

The students' evaluations are represented by 10 evaluation scores--factor scores for the eight evaluation factors already discussed and the Overall Instructor and Overall Course evaluation items. Analysis was performed on both individual student responses and course averages. A random sample of approximately 1,300

individual student responses (with no more than 2 items missing or marked "not applicable") was selected from the entire population of data. The eight factor scores, weighted averages of the evaluations items, were computed for each student; the group mean was substituted for any missing values. Factor scores were also computed for each class. All 10 evaluation scores were standardized (mean 50, standard deviation 10) to make comparison easier. Because of the large sample sizes (1321 randomly selected students or 591 classes), even trivial difference are statistically significant. Relationships accounting for less than 5% of the variance in an evaluation score are dismissed as being unimportant even when statistically significant.

## Results and Discussion

The correlation between expected grades and each of the 10 evaluations is presented in Table One. Although interpretations are based upon class average responses, correlations based upon a randomly selected sample of individual student responses are also presented for purposes of comparison. The findings presented in Table One indicate that expected grades showed substantial correlations with several evaluation scores. Classes of students who, on the average, expected to receive higher grades indicated that their classes were less difficult (accounting for about 22% of the variance in this score), felt their examinations and overall learning experiences were more valuable (accounting for about 14% of the variance in each of these scores), gave the course a higher overall rating (accounting for about 14% of the variance in this score) and gave the instructor a higher

overall evaluation (accounting for about 10% of the variance in this score). The relationship between expected grades and the other five evaluation scores (Enthusiasm, Breadth of Coverage, Organization, Interaction, and Assignments), although statistically significant, was small; expected grades accounted for 4% or less of the variance in any of these scores. The pattern of correlations based upon individual student responses is the same as just described; but without exception, each of these correlations is substantially lower than the same correlation based upon class average responses.

---

Insert Table One about here

---

The low to moderate correlations found here are similar to those found in other studies. Some of the relationships are not surprising and may not have any serious consequences; classes of students expecting to receive lower grades understandably find a course more difficult, are less satisfied with the examinations upon which the grades are based, and, if they believe the lower grades are justified, may feel that they have learned less. However, the substantial correlations between expected grades and the two overall summary items are more serious in that these summary items are often the only ones used to obtain an overall impression of instructional quality .

Overall summary items, being global and non-specific, tend to be more susceptible to response bias. Alan Sockloff (1973, p. 143) contends that the "use of poor, relatively global-type items seems to demand personal response bias rather than objectivity" and suspects "that a good actor who assigns high grades and stimulates little in

the way of learning can fare pretty well on instruments consisting of items that violate most of the guidelines." In support of this contention, it should be noted that even though the four evaluation factors most often used to characterize aspects of teaching (Instructor Enthusiasm, Breadth of Coverage, Interaction and Organization) correlate highly with the overall summary items, they show little relation to expected grades. The obvious conclusion is that if expected grades are a source of bias in students' evaluations, then the factor scores based upon a number of specific items are less biased than the overall summary items. This is particularly important in that many programs of students' evaluations still rely heavily upon these overall summary items rather than on factor scores reflecting distinct components of teaching.

---

Insert Figure 1 about here

---

In order to present a clearer picture of the effect of expected grades upon the 10 evaluation scores, the 591 classes were divided into four groups according to the average grades that each class expected to receive. Mean evaluations for each of the 10 evaluation scores are plotted for the four groups in Figure One. Because the evaluation scores are standardized (mean = 50, standard deviation = 10), the magnitude of the differences between groups is directly comparable to the magnitude of correlations presented in Table One. In addition, the evaluations of 26 "most outstanding" instructors and 26 "least outstanding" instructors are presented to provide a basis of comparison. The previous year, graduating seniors were asked to complete a "Senior Survey" in which, along with other information, they identified the instructor in their major department who had contributed most (and

least) to their educational experience in a classroom setting. The two sets of 26 instructors were identified on the basis of these responses.

While the effects of expected grades are moderate, even classes expecting to receive the highest and the lowest grades are evaluated less extremely than classes taught by the best and worst teachers. The implications are that even if expected grades do bias evaluations--and this is only one possible explanation of the relationship--the bias, even in the most extreme cases is not large. An average instructor who is particularly lenient in assigning grades may be evaluated somewhat better than average (particularly if only overall summary items are considered), but grading leniency will not cause a poor teacher to be evaluated highly or a superior teacher to be evaluated poorly.

BIBLIOGRAPHY

Anikeeff, A. M. Factors affecting student evaluation of college faculty members. Journal of Applied Psychology, 1953, 37, 458-460.

Bausell, R. B. and Magoon, J. Expected grade in a course, grade point average, and student ratings of the course and the instructor. Educational and Psychological Measurement, 1972, 32, 1013-1023.

Caffrey, B. Lack of bias in student evaluations of teachers. Proceedings of the 77th Annual Convention of the American Psychological Association, 1969, 4, 641-642.

Cohen, S. H. and Berger, W. G. Dimensions of students ratings of college instructors underlying subsequent achievement on course examinations. Proceedings of the 78th Annunal Convention of the American Psychologi Association, 1970, 605-606.

Costin, F., Greenough, W. T., and Menges, R. J. Student ratings of college teaching: Reliability, validity, and usefulness. Review of Educational Research, 1971, 41, 511-535.

Dixon, W. J. (Ed.) Biomedical Computer Programs. Berkeley: University of California Press, 1973.

Doyle, K. O. Jr., and Whitely, S. E. Student ratings as criteria for effective teaching. American Educational Research Journal, 1974, 11, 259-274.

Elliot, D. H. Characteristics and relationships of various criteria of college and university teaching. Purdue University Studies in Higher Education, 1950, 70, 5-61.

Frey, P. W. Student ratings of teaching: Validity of several rating factors. Science, 1973, 182, 83-85.

Granzin, K. L. and Painter, J. J. A new explanation for students' course evaluation tendencies. American Educational Research Journal, 1973, 10, 115-124.

Harmon, H. H. Modern Factor Analysis. Chicago: University of Chicago Press, 1967.

Hildebrand, M., Wilson, R. C. and Dienst, E. R. Evaluating University Teaching. Berkeley: Center for Research and Development in Higher Education, University of California, Berkeley, 1971.

Holmes, D. S. Effects of grades and disconfirmed grade expectancies on students' evaluations of their instructor. Journal of Educational Psychology, 1972, 63, 130-133.

Jiobu, R. M. and Pollis, C. A. Student evaluations of courses and instructors. The American Sociologist, 1971, 6, 317-321.

Kooker, E. W. The relationship of known college grades to course ratings on student selected items. The Journal of Psychology, 1968, 69, 209-215.

Marsh, H. W., Fleiner, H., and Thomas, C. S. validity and usefulness of student evaluations of instructional quality, Journal of Educational Psychology, 1975, 67, 833-839.

McKeachie, W. J. Correlates of Student Ratings. In Sockloff, A. L. (Ed.), Proceedings: The First Invitational Conference on Faculty Effectiveness as Evaluated by Students. Philadelphia: Measurement and Research Center, Temple University, 1973.

Morsh, J. E., Burgess, G. G. and Smith, P. N. Student achievement as a measure of instructor effectiveness. Journal of Educational Psychology, 1956, 47, 79-88.

Perry, R. R. and Bauman, R. R. Criteria for evaluation of college teaching: Their reliability and validity at the University of Toledo. In Sockloff, A. L. (Ed.), Proceedings: The First Invitational Conference on Faculty Effectiveness as Evaluated by Students. Philadelphia: Measurement and Research Center, Temple University, 1973.

Remmers, H. H. Teaching methods in research on teaching. In N. L. Gage (Ed.), Handbook of Research on Teaching. Chicago: Rand McNally, 1963.

Remmers, H. H. The relationship between students' marks and students' attitudes toward instructors. School and Society, 1928, 28, 759-760.

Remmers, H. H. To what extent do grades influence student ratings of instructors? Journal of Educational Research, 1930, 21, 314-316.

Roshenshine, B., Cohen, A., and Furst, N. Correlates of student preference ratings. Journal of College Student Personnel, 1973, 14, 269-272.

Sockloff, A. L. Instruments for Student Evaluation of Faculty: Ideal and Actual. In Sockloff, A. L. (Ed.), Proceedings: The First Invitational Conference on Faculty Effectiveness as Evaluated by Students. Philadelphia: Measurement and Research Center, Temple University, 1973.

Staraak, J. A. Student rating of instruction. Journal of Higher Education, 1934, 5, 88-90.

Stewart, C. T. and Malpass, L. F. Estimates of achievement and ratings of instructors. Journal of Educational Research, 1966, 59, 347-350.

Voeks, V. and French, G. M. Are student ratings of teachers affected by grades? Journal of Higher Education, 1960, 31, 330-334.

Weigel, R. G., Oetting E. R. and Tasto, D. L. Differences in course grades and student ratings of teacher performance. School and Society, 1971, 99, 60-62.

## Footnotes

[1]Requests for reprints should be sent Herbert W. Marsh, President, Evaluation, Testing and Research, Inc., 1110 Lake St., Suite #3, Venice, California 90291

[2]The evaluation instrument described in this research was the basis for the commercially available Student Evaluation of Education (SEE) instrument. Inquiries should be sent to the first author of this article or to Evaluation, Testing & Research.

Table One

Correlations Between Expected Grades and 10 Evaluation Scores[1]

| Evaluation Scores | Class Average Responses (n = 591 classes) | Individual Student Responses (n = 1321 students) |
|---|---|---|
| Instructor Enthusiasm | .13 ** | .03 |
| Breadth | .14 ** | .02 |
| Organization | .12 * | .03 |
| Interaction | .18 *** | .08 ** |
| Learning | .38 *** | .21 *** |
| Exams | .38 *** | .21 *** |
| Assignments | .21 *** | .16 *** |
| Difficulty | .47 *** | .30 *** |
| Overall Instructor | .32 *** | .13 *** |
| Overall Course | .37 *** | .17 *** |

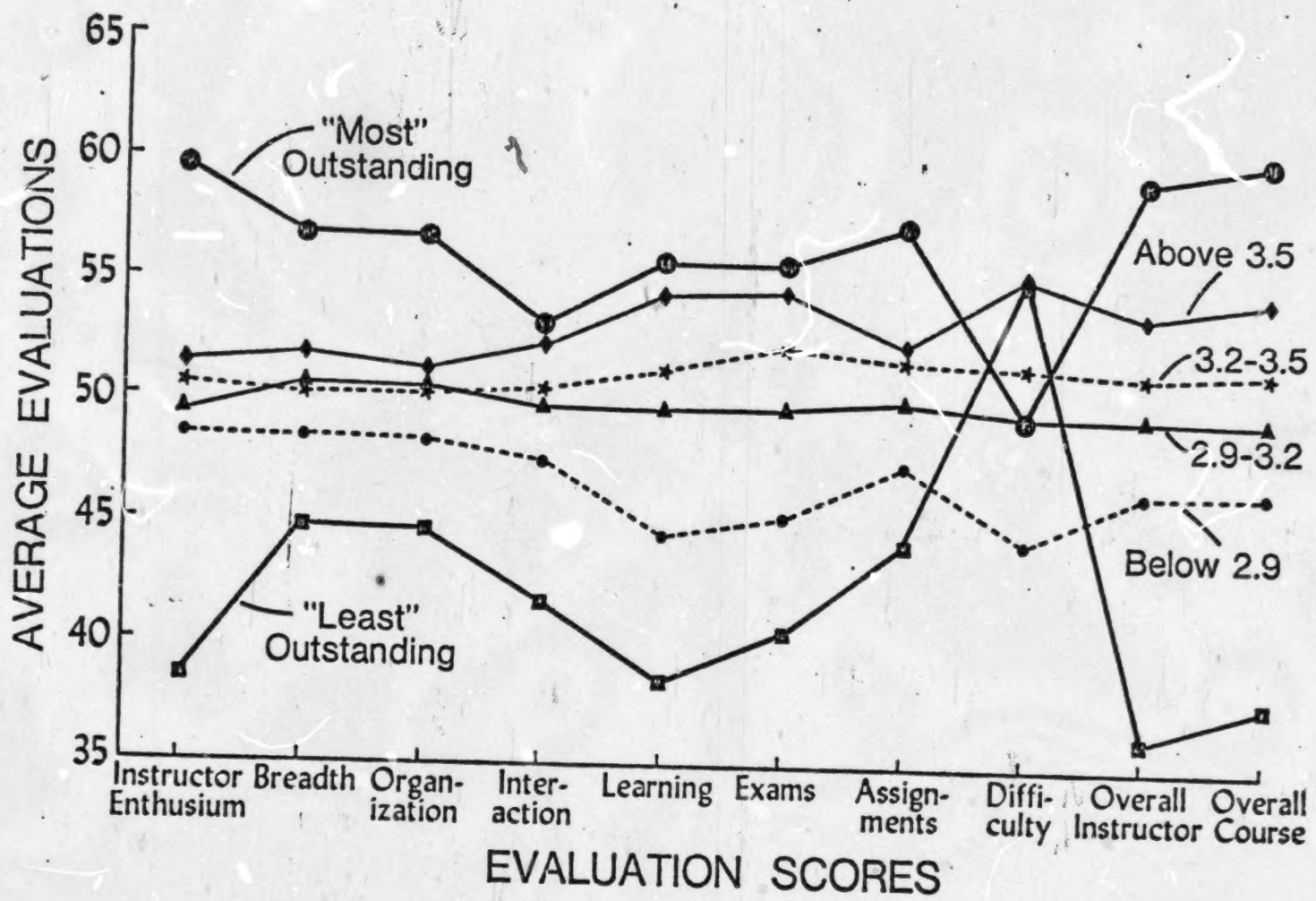
* p.=.05, ** p.=.01, *** p.=.001

[1]The direction of the correlations indicate that higher expected grades are associated with higher evaluations and less difficult courses.

## Figure Caption

Figure One. Class average evaluation scores for courses differing in average grades expected by students as compared to class average evaluation scores of courses taught by instructors who were independently identified as good (most outstanding) and poor (least outstanding) teachers. ("Difficulty" scores have been reversed so that higher scores reflect easier courses.)

FIGURE ONE


AVERAGE EVALUATIONS
65
60
55
50
45
40
35
"Most" Outstanding
"Least" Outstanding
Above 3.5
3.2-3.5
2.9-3.2
Below 2.9
Instructor Enthusium
Breadth
Organ-ization
Inter-action
Learning
Exams
Assign-ments
Diffi-culty
Overall Instructor
Overall Course
EVALUATION SCORES